Dynamics of Scale-Free Semi-Synchronous Boolean Networks

M. Giacobini[1], M. Tomassini[1], P. De Los Rios[2], and E. Pestelacci[1]
[1]Information Systems Institute, University of Lausanne, Lausanne, Switzerland
[2]Laboratory of Statistical Biophysics, Ecole Polytechnique Federale de Lausanne - EPFL, Lausanne, Switzerland
{mario.giacobini,marco.tomassini,enea.pestelacci}@unil.ch, paolo.delosrios@epfl.ch

Abstract

Random Boolean Networks have been introduced by Kauffman more than thirty years ago as a highly simplified model of genetic regulatory networks. These models are interesting in their own as complex dynamical systems and have been throughly studied as such. We believe that the original view of Kauffman is still a valid one, provided that the model is updated to take into account present knowledge, without loosing its attractive simplicity. Thus, we will present how the Kauffman model could be modified in order to qualitatively agree with experimental observations that were not available at the time. In particular, we will present a method for generating networks with given degree distributions, together with a new semi-synchronous updating scheme. Simulations of statistical ensembles of networks behaving according to the new model will be presented and discussed.

Introduction

Random Boolean Networks (RBN) have been introduced by Kauffman more than thirty years ago in a landmark paper (Kauffman, 1969) as a highly simplified model of genetic regulatory networks. In a RBN with N nodes, a node represents a gene and is modeled as an on-off device, meaning that a gene is expressed if it is on (1), and it is not otherwise (0). Each gene receives K randomly chosen inputs from other genes. Initially, one of the possible Boolean functions of K inputs is assigned at random to each gene. The network dynamics is discrete and synchronous: at each time step all nodes simultaneously examine their inputs, evaluate their Boolean functions, and find themselves in their new states at the next time step. Over time, the system travels through its phase space, until a point or cyclic attractor is reached whence either it will remain in that point attractor forever, or it will cycle through the states of the periodic attractor. Since the system is finite and deterministic, this will happen at most after 2^N time steps.

This extremely simple and abstract model has been studied in detail by analysis and by computer simulations of statistical ensembles of networks and it has been shown to be capable of extremely interesting dynamical behavior. Complete descriptions can be found in (Kauffman, 1993; Aldana et al., 2003). We summarize the main results here.

First of all, it has been found that, as some parameters are varied such as K, or the probability p of expressing a gene, i.e. of switching on the corresponding node's state, the RBN can go through a phase transition. Indeed, for every value of p, there is a critical value of connectivity K such that for values of K below this critical value the system is in the ordered regime, while for values of K above this limit the system is said to be in the chaotic regime. In classical RBN $K = 1$ corresponds to the ordered regime, $K = 2$ is critical, and $K \geq 3$ means that the system is in the chaotic phase. Kauffman found that for $K = 2$ the size distribution of perturbations in the networks is a power law with finite cutoff that scales as the square root of N. Thus perturbations remain localized and do not percolate through the system. The mean cycle length scales at most linearly with N for $K = 2$. Kauffman's suggestion is that cell types correspond to attractors in the RBN phase space, and only those attractors that are short and stable under perturbations will be of biological interest. Thus, according to Kauffman, $K = 2$ RBN lying at the edge between the ordered phase and the chaotic phase can be seen as abstract models of genetic regulatory networks.

RBN are interesting in their own as complex dynamical systems and have been throughly studied as such using the concepts and tools of statistical mechanics (see (Derrida and Pomeau, 1986; Aldana et al., 2003)). There is nothing wrong with this; however, we believe that the original view of Kauffman, namely that these models may be useful for understanding real cell regulatory networks, is still a valid one, provided that the model is updated to take into account present knowledge about the topology of real gene regulatory networks, and the timing of events, without loosing its attractive simplicity.

In the following sections we shall describe how, in our opinion, the Kauffman model could be modified in order to take into account a number of experimental observations that were not available at the time. We shall then describe and comment the results of the simulations of statistical ensembles of networks behaving according to the new model. Finally, we shall present our conclusions as well as some ideas for future studies.

The Network Model

Kauffman's RBN model rests on three main assumptions:

- The nodes implement Boolean functions and their state is either on or off;
- The nodes that affect a given node in the network are randomly chosen and are a fixed number;
- The dynamics of the network is synchronous in time.

The binary state simplification could seem extreme but actually it represents quite well "threshold phenomena" in which variables of interest suddenly change their state, such as neurons firing or genes being switched on or off.

Random networks with fixed connectivity degree were a logical generic choice in the beginning, since the exact couplings in networks were generally unknown. Today it is more open to criticism since it does not correspond to what we know about the topology of biological networks. In fact, many biological networks, including genetic regulatory networks, seem to be of the scale-free type or of a hierarchical type (see (Vázquez et al., 2004) and references therein) but not random, according to present data. For scale-free networks, this means that the distribution function of the degree, i.e. the probability $P(k)$ that a randomly chosen node has degree k, is a power law $P(k) \sim k^{-\gamma}$, usually with $2 < \gamma < 3$, instead of a Poisson distribution as in a random graph, or a delta distribution as in a classical RBN. Thus the low connectivity suggested by Kauffman for candidate stable systems is not found in such networks, where a wide range of degrees is present instead. The consequences for the dynamics may be important, since in scale-free graphs there are many nodes with low degree and a low, but not vanishing, number of highly connected nodes (see, for instance, the review (Albert and Barabasi, 2002)).

The first work that we are aware of using the scale-free topology for Boolean networks dynamics is (Oosawa and Savageau, 2002). Oosawa and Savageau took *Escherichia coli* as a model for their scale-free nets with an average input degree $\overline{K}$ of two. But, although interesting in this particular case, this is too limited as most other networks have higher connectivity levels. What is needed are models that span the range of observed connectivities.

Along this line, M. Aldana has recently presented a detailed analysis of Boolean networks with scale-free topology (Aldana, 2003). He has been able to define a phase space diagram for boolean networks, including the phase transition from ordered to chaotic dynamics, as a function of the power law exponent γ. He also made exhaustive simulations for several relatively small values of N, the network size.

Our model has in common with Aldana's the scale-free topology of the networks, although the graphs are constructed in a different way. But, in contrast to Aldana's, we define a suitable asynchronous dynamics for the system, instead of using the customary synchronous update. We motivate our choices and describe the model below. Since Aldana's result are the first systematic data available for scale-free Boolean networks, we shall compare our results with his, and with the standard model in the sequel.

Network Construction

Kauffman's RBN, also known as $N-K$ models, are directed graphs. In fact, if a node receives an input from another node in the graph, this does not necessarily mean that the latter receives an input from the former node. Let's suppose that each node i ($i \in \{1,\ldots,N\}$) receives k_i inputs and projects a link to other l_i nodes, i.e. there are l_i nodes in the graphs that receive an input from node i. Among the N nodes of the graph, the distribution $P_{in}(k)$ of the input connections is not necessarily the same of the distribution of the output connections $P_{out}(k)$.

According to present data, many biological networks, including genetic regulatory networks, show a scale-free output distribution $P_{out}(k)$ and a Poissonian input distribution $P_{in}(k)$ (Vázquez et al., 2004). The *preferential attachment* rule described in (Albert and Barabasi, 2002) could be easily modified to construct networks with such input and output distributions, simply interpreting any newly added edge as a directed edge impinging on the added node. This method would produce only a subfamily of the graphs sharing these degree distributions, and, in particular, does not permit to generate networks having scale-free distributions of the output degrees with different exponents.

In order to avoid such a bias towards a particular subfamily of graphs, the networks used for the present investigation have been generated according to a mixed generalized/poisson random graph model (Bender and Canfield, 1978; Molloy and Reed, 1995): first a sequence of N out-degrees that satisfies a power-law distribution with exponent γ is assigned to N nodes; then, every out-going edge is assigned as input to one of the N nodes chosen at random (excluding self-connections). The resulting networks have a scale-free distribution of the output degrees and a Poisson distribution of the input degrees.

Synchronous, Asynchronous and Semi-Synchronous Network Dynamics

Standard RBN update their state synchronously. This assumption simplifies the analysis, but it is open to discussion if the network has to be biologically plausible. In particular, for genetic regulatory networks, this is certainly not the case, as many recent experimental observations tend to prove. Rather, genes seem to be expressed in different parts of the network at different times, according to a strict sequence (see, for instance, (Davidson and et al., 2002)). Thus a kind of serial, asynchronous update sequence seems to be needed. Asynchronous dynamics must nevertheless be fur-

ther qualified, since there are many ways for serially updating the nodes of the network.

Two types of asynchronous updates are commonly used. In the first, a random permutation of the nodes is drawn and the nodes are updated one at a time in that order. At the next update cycle, a fresh permutation is drawn and the cycle repeated. Let us call this policy *Random Permutation Update* (RPU). In the second policy, the next cell to be updated is chosen at random with uniform probability and with replacement. This is a good approximation of a continuous-time Poisson process, and it will be called *Uniform Update* (UU).

Several researchers have investigated the effect of asynchronous updating on classical RBN dynamics in recent years (Harvey and Bossomaier, 1997; Mesot and Teuscher, 2003; Rohlfshagen and Di Paolo, 2004; Gershenson, 2002; Gershenson, 2004). Harvey and Bossomayer studied the effect of random asynchronous updating on some statistical properties of network ensembles, such as cycle length and number of cycles, using both RPU and UU (Harvey and Bossomaier, 1997). They found that many features that arise in synchronous RBN do not exist, or are different in non-deterministic asynchronous RBN. Thus, while point attractors do persist, there are no true cyclic attractors, only so-called loose ones and states can be in more than one basin of attraction. Also, the average number of attractors is very different from the synchronous case: even for $K = 2$ or $K = 3$, which are the values that characterize systems at the edge of chaos, there is no correspondence between the two dynamics.

Mesot and Teuscher studied the critical behavior of asynchronous RBN and concluded that they do not have a critical connectivity value analogous to synchronous RBN and they behave, in general, very differently from the latter, thus confirming in another way the findings of (Harvey and Bossomaier, 1997).

Rohlfshagen and Di Paolo (Di Paolo E. A., 2001; Rohlfshagen and Di Paolo, 2004) mainly investigated rythmic and non-rythmic attractors using evolutionary algorithms. They showed that, in spite of non-determinism, asynchronous RBN can be evolved that possess core structures with circular topology and can model coordinated rythmic behavior.

Gershenson (Gershenson, 2002; Gershenson, 2004) extended the analysis and simulation of asynchronous RBN by introducing additional update policies in which specific groups of nodes are updated deterministically. He found that all types of networks have the same point attractors but other properties, such as the size of the attractor basins and the cyclic attractors do change. One of Gershenson's conclusions (Gershenson, 2004) is that the main differences are due more to non-determinism of the update than to asynchronicity and that classical synchronous RBN are plausible models for theoretical studies.

Considering the above reults and what is known experimentally about the timing of events in genetic networks we conclude, with (Mesot and Teuscher, 2003), that neither fully synchronous nor completely random asynchronous network dynamics are suitable models. Synchronous update is implausible because events do not happen all at once, while completely random dynamics does not agree with experimental data on gene activation sequences and the model does not show stable cyclic attractors of the right size.

In genetic regulatory networks, the expression of a gene depends on some transcription factors, whose synthesis does not appear to be neither fully synchronous nor instantaneous. Moreover, in some cases like the gene regulatory network controlling embryonic specification in the sea urchin (Davidson and et al., 2002; Olivieri and Davidson, 2004), we can clearly see the presence of an activation sequence of genes. In our opinion, the activation/update sequence in a RBN should be in some way related to the topology of the network.

In this paper we propose a new topology-driven semi-synchronous update method, called *Cascade Update* (CU). However, we do not claim that such an update scheme is a faithful model for true biological gene activation sequences which, on the other hand, are clearly not the same for different regulatory networks. But we do believe that our proposed scheme is closer to biological reality than the previously proposed ones namely, fully synchronous and various asynchronous policies.

The CU consists in an asynchronous sequence of synchronously updated blocks of nodes. At the beginning of the evolution of the network, a node, say *i*, is randomly chosen and updated. Then, in the next time step, the block of the nodes to whom *i* projects is synchronously updated. The process continues, updating at each time step a new block formed by all the nodes in the network to which the nodes updated in the previous time step project. This scheme is deterministic: once the first node is chosen, the sequence of all the successive updates is unique and will reach a cycle, since the dynamical system is finite. As a consequence, the attractors of the dynamics cannot be loose attractors, they have to be true point or cyclic ones.

The behavior of this new semi-synchronous update method will be compared to the synchronous and the asynchronous UU ones in the next section.

Simulation Methodology

The method to generate the networks described in the previous section permit to produce directed graphs with a scale-free distribution of the output degrees and a Poisson distribution of the input degrees. In particular, the γ exponent of the scale-free distribution can be fixed to a given value, which is, according to (Aldana, 2003), $\gamma \in [2, 2.5]$ for the majority of the biological networks analyzed up to now. Moreover, he investigated the phase diagram of scale-free $N-K$ model, showing that the critical value of the γ exponent that discrim-

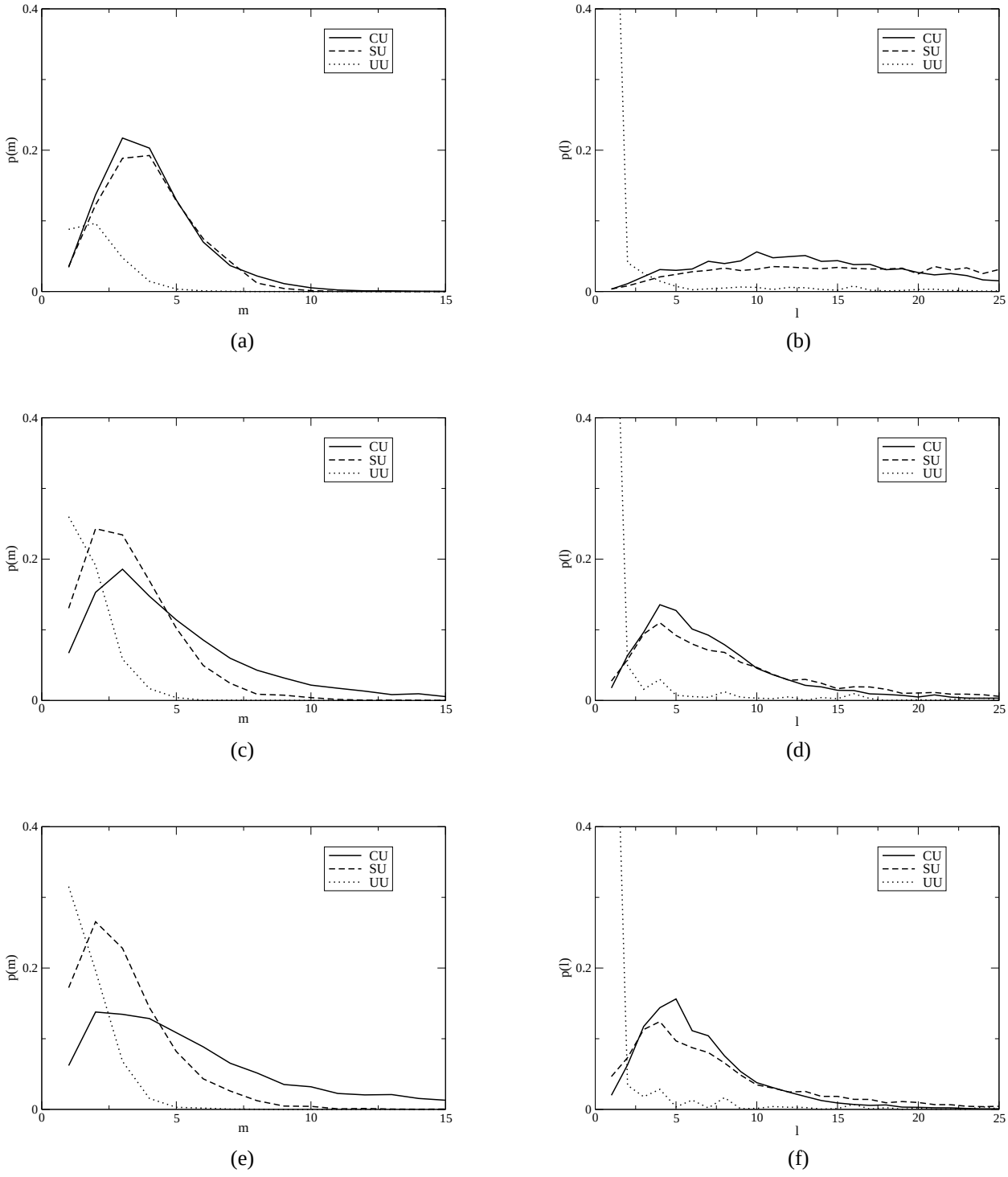


Figure 1: Simulation results of network realizations having $N = 19$ nodes with three different scale-free output degree distribution exponents: $\gamma_1 = 1.1$, i.e. chaotic phase (probability distribution of the number of different attractors (a), and probability distribution of the length of the attractors (b)), $\gamma_2 = 2.48$, i.e. edge of chaos (probability distribution of the number of different attractors (c), and probability distribution of the length of the attractors (d)), and $\gamma_3 = 3$, i.e. ordered phase (probability distribution of the number of different attractors (e), and probability distribution of the length of the attractors (f)).

inates the ordered regime from the chaotic phase when the probability p of gene expression is equal to 0.5 is $\gamma_c = 2.48$.

Following Aldana's approach, we have decided to analyze the different behaviors showed by the synchronous, the asynchronous UU and the new semi-synchronous CU on scale-free RBN with three γ exponents. The three values are chosen to be $\gamma_1 = 1.1$ for the chaotic phase, $\gamma_2 = 2.48$ for the *edge of chaos*, and $\gamma_3 = 3$ for the ordered regime.

Analysis of the Results

In order to compare our results with Aldana's, we first investigated the dynamical behavior of networks with $N = 19$ nodes. For each of the 200 graph realizations (each one being an instance of a directed graph with a scale-free distribution of the output degrees and a Poisson distribution of the input degrees), 20 sets of node functions are randomly generated with a fixed probability $p = 0.5$ of expressing the gene, thus obtaining 4000 network realizations. The experimental results presented here were computed by randomly sampling 500 initial configurations for each network realization, and then evolving the system for $t = 20000$ time steps. For the simulations we have used a slightly modified version of Gershenson's RBNLab software (http://rbn.sourceforge.net).

For the ensemble of these evolution we monitored the probability $p(m)$ for a network realization to have exactly m different attractors, and the probability $p(l)$ of an attractor to contain l different states. Figure 1 shows the two statistics for network realizations in the chaotic phase ($\gamma_1 = 1.1$), the critical or edge of chaos border ($\gamma_2 = 2.48$), and the ordered regime ($\gamma_3 = 3$).

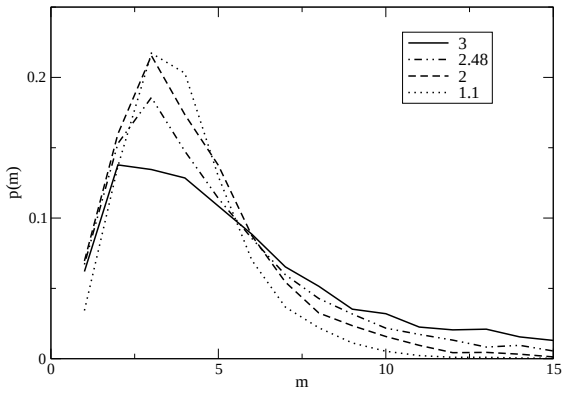
The networks with $\gamma_3 = 3$ (Figure 1 (e)) shows that the synchronous and the semi-synchronous updates have an high probability of having more than one attractor. In the ordered phase, synchronous RBN are more likely to have 2 or 3 attractors, while CU RBN have 2, 3, 4, or 5 attractors with about the same probability. These attractors have a size going from 3 to 7 states per attractors (Figure 1 (f)). Attractors with size greater than 10 are also possible.

The networks with $\gamma_2 = 2.48$ (Figure 1 (c)–(d)) show similar probability distributions of the quantity and of the lengths of the attractors for each network realization. The networks updated with the CU scheme show more attractors and these attractors are larger.

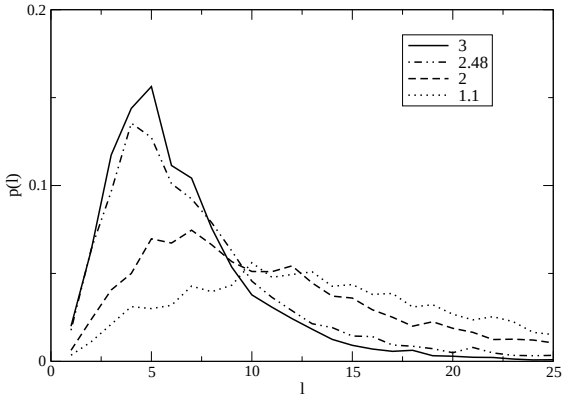
In networks with $\gamma_1 = 1.1$ (Figures 1 (a)) we can see that the networks always have more attractors if updated with a synchronous or semi-synchronous scheme. Moreover, the difference between the two schemes becomes very small. Similar observations can be made on the size of the attractors (Figures 1 (b)).

Finally, the non-deterministic fully asynchronous UU update does not show biologically interesting behaviors, since the attractor are mostly fixed points or very short length ones. This confirms the findings of (Harvey and Bossomaier, 1997; Gershenson, 2002).

Focusing only on the semi-synchronous CU, Figure 2 depicts the probability distributions of the number of attractors $p(m)$ per network realization, and of the length of an attractor $p(l)$, with γ exponent in the set $\{1.1, 2, 2.48, 3\}$. A smaller exponent implies less attractors, with the edge of chaos distribution clearly staying between the chaotic and the ordered phase distributions.



(a)



(b)

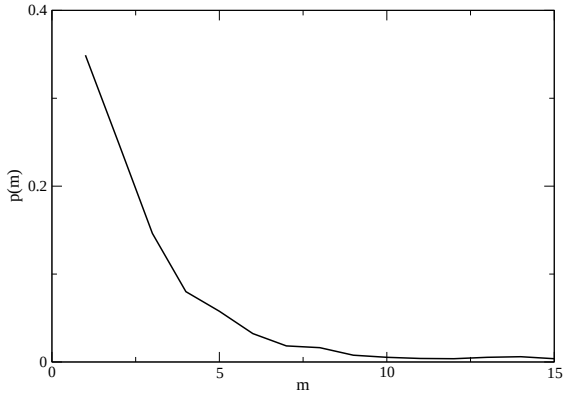
Figure 2: Comparison of the probability distributions of the number of different attractors (a), and of the length of the attractors (b) for network realizations having $N = 19$ nodes evolving using the CU scheme when varying the γ exponent in the set $\{1.1, 2, 2.48, 3\}$.

Larger Networks

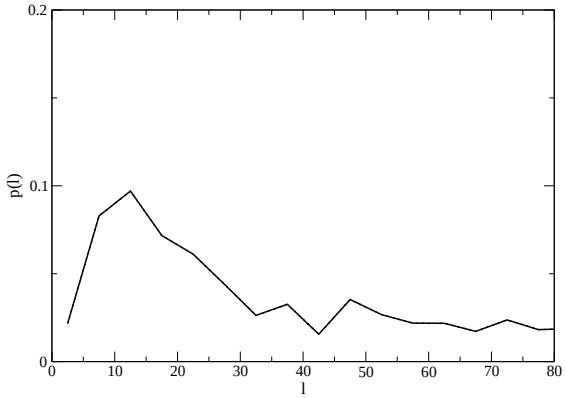
In a second simulation phase we have analyzed the behavior of larger scale-free RBN with dynamics dictated by the semi-synchronous node update CU. We have generated network realizations with $N = 50$ nodes. These networks have

a more reasonable size if compared to the known genetic regulatory networks. For example, the GRN that controls the embryonic specification in the sea urchin shows a connected component of 47 genes (see (Davidson and et al., 2002; Olivieri and Davidson, 2004)).

We only simulated networks with γ exponent equal to 2.48. Figure 3 depicts the probability distributions of the number of attractors $p(m)$ per network realization, and of the length of an attractor $p(l)$. The large majority of these network realizations, around 50%, show 2 to 5 different attractors, which is biologically reasonable. These attractors are larger in size with respect to those found in the first simulation phase.



(a)



(b)

Figure 3: probability distributions of the number of different attractors (a), and of the length of the attractors (b) for network realizations having $N = 50$ nodes evolving using the CU.

Conclusions and Future Work

In this study we have presented a new semi-synchronous model for the dynamics of scale-free RBN that is more respectful of known present-day genetic regulatory network data. We have experimentally shown that the model behavior is close to Aldana's synchronous dynamics (Aldana, 2003) for scale-free RBN. However, in our case the dynamics is driven by a closer approximation to the timing of events in observed biological regulatory networks.

The attractor structure of the new phase space is biologically plausible, although we cannot hope, of course, to be faithful with respect to actual genetic regulatory networks with such a simple and abstract model. Nevertheless, the present study confirms the interest of working with simplified models *à la Kauffman*, provided that the influence of the network topology on the dynamics is taken into account so that actual gene expression patterns are roughly reproduced.

Our current work is focused on the study of the semi-synchronous model's behavior on network topologies that are built by taking into account data from the recent biological literature (Davidson and et al., 2002; Olivieri and Davidson, 2004). Also, since we assumed Aldana's parameters for characterizing the main features of the system's phase space, we plan to investigate in more detail the phase space structure with the new dynamics, extending the study to statistical ensembles of larger networks.

Acknowledgements

The authors would like to thank F. Di Cunto and P. Provero of the University of Turin (Italy) for the useful discussions and suggestions on biological regulatory networks.

M. Tomassini gratefully acknowledges financial support by the Swiss National Science Foundation under contract 200021-107419/1.

References

Albert, R. and Barabasi, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97.

Aldana, M. (2003). Boolean dynamics of networks with scale-free topology. *Physica D*, 185:45–66.

Aldana, M., Coppersmith, S., and Kadanoff, L. P. (2003). Boolean dynamics with random couplings. In Kaplan, E., Marsden, J. E., and Sreenivasan, K. R., editors, *Perspectives and Problems in Nonlinear Science*, Springer Applied Mathematical Sciences Series, pages 23–89. Springer, Berlin.

Bender, E. and Canfield, E. (1978). The asymptotic number of labeled graphs with given degree distribution. *Journal of Combinatorial Theory A*, 24:296–307.

Davidson, E. H. and et al. (2002). A genomic regulatory network for development. *Science*, 295:1669–1678.

Derrida, B. and Pomeau, Y. (1986). Random networks of automata: a simple annealed approximation. *Europhysics Letters*, 1(2):45–49.

Di Paolo E. A. (2001). Rythmic and non-rythmic attractors in asynchronous random Boolean networks. *BioSystems*, 59:185–195.

Gershenson, C. (2002). Classification of random Boolean networks. In Standish, R., editor, *Artificial Life VIII: Proceedings of the Eight International Conference on Artificial Life*, pages 1–8. MIT Press.

Gershenson, C. (2004). Updating schemes in random Boolean networks: Do they really matter? In Pollack, J., editor, *Artificial Life IX Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*, pages 238–243. MIT Press.

Harvey, I. and Bossomaier, T. (1997). Time out of joint: attractors in asynchronous random boolean networks. In Husbands, P. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life*, pages 67–75, Cambridge, MA. The MIT Press.

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, 22:437–467.

Kauffman, S. A. (1993). *The Origins of Order*. Oxford University Press, New York.

Mesot, B. and Teuscher, C. (2003). Critical values in asynchronous random boolean networks. In Banzhaf, W., editor, *Advances in Artificial Life, ECAL2003*, volume 2801 of *Lecture Notes in Artificial Intelligence*, pages 367–376, Berlin. Springer.

Molloy, M. and Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random. Struct. Algorithms*, 6:161.

Olivieri, P. and Davidson, E. H. (2004). Genes regulatory network controling embryonic specification in the sea urchin. *Current Opinion in Genetics and Development*, 14:351–360.

Oosawa, C. and Savageau, M. A. (2002). Effects of alternative connectivity on behavior of randomly constructed Boolean networks. *Physica D*, 170:143–161.

Rohlfshagen, P. and Di Paolo, E. A. (2004). The circular topology of rythm in asynchronous random Boolean networks. *BioSystems*, 73:141–152.

Vázquez, A., Dobrin, R., Sergi, D., Eckmann, J.-P., Oltvai, Z. N., and Barabàsi, A.-L. (2004). The topological relationships between the large-scale attributes and local interactions patterns of complex networks. *Proc. Natl. Acad. Sci USA*, 101(52):17940–17945.